Whitney SM, Birch R, Kelso C, Beck JL, Kapralov MV.

Improving recombinant Rubisco biogenesis, plant photosynthesis and growth by coexpressing its ancillary RAF1 chaperone.

***Proceedings of the National Academy of Sciences of the United States of America* 2015, 112(11), 3564–3569**

Classification: Biological Sciences –Plant Biology

Improving recombinant Rubisco biogenesis, plant photosynthesis and growth by co-expressing its ancillary RAF1 chaperone

Enhancing Rubisco assembly by RAF1 co-expression

Spencer M. Whitney*[‡], Rosemary Birch*, Celine Kelso[†], Jennifer L. Beck[†], Maxim V. Kapralov*

* Research School of Biology, The Australian National University, Acton, Australian Capital Territory 2601, Australia. [†]School of Chemistry, University of Wollongong, New South Wales 2522, Australia

[‡] Corresponding author; Research School of Biology, Australian National University, Acton, Australian Capital Territory 2601, Australia; Tel: +61-2-6125-5073; E-mail: spencer.whitney@anu.edu.au

Number of text pages: 29
Number of figures: 5
Number of Tables: 0
Number of characters (including spaces): 42,135

Abstract

Enabling improvements to crop yield and resource use by enhancing the catalysis of the photosynthetic CO_2-fixing enzyme Rubisco has been a longstanding challenge. Efforts towards realization of this goal have been greatly assisted by advances in understanding the complexities of Rubisco's biogenesis in plastids and the development of tailored chloroplast transformation tools. Here we generate transplastomic tobacco genotypes expressing *Arabidopsis* Rubisco large subunits (AtL) both on their own (producing tob^{AtL} plants) and with a cognate Rubisco Accumulation Factor 1 (AtRAF1) chaperone (producing $tob^{AtL\text{-}R1}$ plants) that has undergone parallel functional co-evolution with AtL. We show AtRAF1 assembles as a dimer and is produced in $tob^{AtL\text{-}R1}$ and *Arabidopsis* leaves at 10 to 15 nmol AtRAF1 monomers per m^2. Consistent with a post-chaperonin L-subunit assembly role, the AtRAF1 facilitated two to three fold improvements in the amount and biogenesis rate of hybrid $L_8{}^{A}S_8{}^{t}$ Rubisco (comprising AtL and tobacco small (S) subunits) in $tob^{AtL\text{-}R1}$ leaves compared to tob^{AtL}, despite >3-fold lower steady state Rubisco mRNA levels in $tob^{AtL\text{-}R1}$. Accompanying 2-fold increases in photosynthetic CO_2-assimilation rate and plant growth were measured for $tob^{AtL\text{-}R1}$ lines. These findings highlight the importance of ancillary protein complementarity during Rubisco biogenesis in plastids, the possible constraints this has imposed on Rubisco adaptive evolution and the likely need for such interaction specificity to be considered when optimizing recombinant Rubisco bioengineering in plants.

Significance statement

Using a translational photosynthesis approach we successfully increased CO_2-assimilation in leaf chloroplasts of the model plant tobacco. Phylogenetic analysis revealed parallel evolutionary linkages between the large (L-) subunit of the CO_2-fixing enzyme Rubisco and its molecular chaperone RAF1. We experimentally test, and exploit, this correlation using plastome transformation producing plants that demonstrate the role of RAF1 in L-subunit assembly and resolve the RAF1 quaternary structure as a dimer. We show the increase in Rubisco biogenesis translated to improvements in leaf photosynthesis and growth of the plants. The outcomes have application to the growing interest into identifying, and implementing, strategies to supercharge photosynthesis to improve crop productivity and stem global food security concerns.

\body

Introduction

The increasing global demands for food supply, bioenergy production and CO_2-sequestration have placed a high need on improving agriculture yields and resource use (1, 2). It is now widely recognized that yield increases are possible by enhancing the light harvesting and CO_2-fixation processes of photosynthesis (3-5). A major target for improvement is the enzyme Rubisco (ribulose-1,5-bisphosphate (RuBP) carboxylase/oxygenase) whose deficiencies in CO_2–fixing speed and efficiency pose a key limitation to photosynthetic CO_2 capture (6, 7). In plants, the complex, multistep catalytic mechanism of Rubisco to bind its 5-carbon substrate RuBP, orient its C-2 for carboxylation, and then process the 6-carbon product into two 3-phosphoglycerate (3PGA) products, limits its throughput to 1-4 catalytic cycles per second (8). The mechanism also makes Rubisco prone to competitive inhibition by O_2 that produces only one 3PGA and 2-phosphoglycolate (2PG). Metabolic recycling of 2PG by photorespiration requires energy and results in most plants losing 30% of their fixed CO_2 (5). To compensate for these catalytic limitations plants like rice and wheat invest up to 50% of the leaf protein into Rubisco which accounts for ~25% of their leaf nitrogen (9).

Natural diversity in Rubisco catalysis demonstrates that plant Rubisco is not the pinnacle of evolution (6, 7). Better performing versions in some red algae have the potential to raise the yield of crops like rice and wheat by as much as 30% (10). Bioengineering Rubisco in leaves therefore faces two key challenges: identifying the structural changes that promote performance and identifying ways to efficiently

transplant these changes into Rubisco within a target plant. A significant hurdle to both challenges is the complex biogenesis requirements of Rubisco in plant chloroplasts (7, 11). A number of ancillary proteins are required to correctly process and assemble the chloroplast made Rubisco large (L) subunit (coded by the plastome *rbc*L gene) and cytosol made small (S) subunits (coded by multiple *Rbc*S genes in the nucleus) into L_8S_8 complexes in the chloroplast stroma. The complicated assembly requirements of Rubisco in chloroplasts prevent their functional testing in *E. coli* and conversely impedes, sometimes prevents, the biogenesis of Rubisco from other higher plants, cyanobacteria and algae (12-14). For example, the L-subunits from sunflower and varying *Flaveria* sp. showed 5-fold differences in their capacity to form hybrid L_8S_8 Rubisco (that comprise tobacco S-subunits) in tobacco chloroplasts despite each *rbc*L transgene sharing the same genetic regulatory sequences and showing >92% amino acid identity (13, 14). Evidently evolution of Rubisco function may have been constrained to maintain compatibility with the molecular chaperones required for its biogenesis (7, 15).

The necessity of chloroplast chaperonin (CPN) complexes for Rubisco biogenesis has been known for some time (16). Upon release from the hetero-oligomeric CPN ring structures in chloroplasts (17) the folded L-subunits are thought to sequentially assemble into dimers (L_2) then octamers $(L_2)_4$ prior to S-subunit binding (18). The molecular details of this process remain unclear. The maize Photosynthetic Mutant Library has provided useful insight by identifying three chaperones with roles associated with Rubisco synthesis, assembly and/or stability: Rubisco accumulation factors 1 (RAF1,(19)) and 2 (RAF2; a Pterin-4a-Carbinolamine Dehydratase-like protein, (20))

and BSDII (a DnaJ-like protein, (21)). Results of chemical crosslinking experiments in maize leaves suggest all three proteins might associate with the S-subunit during Rubisco biogenesis (20). Other studies however suggest RAF1 interacts with post-CPN folded L-subunits to assist in L_2 then $(L_2)_4$ formation (19, 22). This function mirrors that shown for RbcX, a Rubisco chaperone that acts as a 'molecular staple' to assemble folded L-subunits into L_2 units for $(L_2)_4$ assembly prior to S-subunit binding to displace the RbcX and trigger catalytic potential (18). While the function of RbcX in L_8S_8 Rubisco biogenesis has been resolved in exquisite molecular detail *in vitro* and in *E. coli*, its functional role in cyanobacteria and in leaf chloroplasts remain unresolved. Comparable molecular details on RAF1, RAF2 and BSDII structure and function remain incomplete, making it difficult to reliably assign their roles and interactions with Rubisco in chloroplasts.

Targeted transformation of the chloroplast genome (plastome) provides a reliable, but time consuming, tool for engineering Rubisco (23). This technology is best developed in tobacco with the ^{cm}trL genotype specifically made for bioengineering Rubisco and testing its effects on leaf photosynthesis and growth (6, 7, 13, 14). Here we use chloroplast transformation in ^{cm}trL to examine the function of RAF1 from *Arabidopsis* (AtRAF1) in Rubisco biogenesis. We show that AtRAF1 forms a stable dimer that, when co-expressed with its cognate *Arabidopsis* Rubisco L-subunits (AtL), enhances hybrid $L_8{}^{A}S_8{}^{t}$ Rubisco (containing *Arabidopsis* L- and tobacco S-subunits) assembly in tobacco chloroplasts and concomitantly improves leaf photosynthesis and plant growth by more than 2-fold.

Results

Co-evolution of RAF1 and the Rubisco L-subunit

Analysis of full length *raf1* and *rbc*L sequences from plant, algae and cyanobacteria showed that Rubisco L-subunit and RAF1 phylogenies are topologically similar (Fig. 1A). Mirror-tree analysis revealed that the correlation coefficient of these trees was 0.75 ($p < 10^{-6}$) suggesting co-evolution of both proteins across cyanobacteria and plants (Fig S1). Exceptionally high correlations between RAF1 and Rubisco L-subunit pairwise non-synonymous distances (i.e. those leading to amino acid substitutions) across all the taxa confirmed co-evolution of the two proteins (Fig 1B). We therefore sought to test the functional significance of this complementarity by transforming the *Arabidopsis* Rubisco L-subunit (AtL) and one of its two cognate RAF1 isoforms (called AtRAF1; Fig S1) into tobacco chloroplasts via plastome transformation. Based on our previous heterologous Rubisco expression studies in tobacco (13, 14) we hypothesized that the phylogenetic divergence of AtL and the tobacco L-subunits (tobL, Fig 1A) would be accompanied by differences in ancillary protein requirements that would impede the biogenesis of hybrid $L_8^{A}S_8^{t}$ Rubisco (*i.e.* comprising AtL and tobacco S-subunits) in tobacco chloroplasts.

Plastome transformation of *Arabidopsis* Rubisco AtL-subunits and AtRAF1 into tobacco chloroplasts

The L-subunit of *Arabidopsis* shares 94% identity with tobL, differing by only 29 amino acids (Fig S2A). Transplanting the *Arabidopsis rbc*L gene (At*rbc*L) into the tobacco plastome in place of the native *rbc*L gene was achieved by cloning it into the plastome-transforming plasmid pLEV4 to give plasmid pLEVAtL and transforming it into the

plastome of the ^{cm}trL tobacco genotype to produce tob^{AtL} lines (Figure 2A). To test the influence of co-expressing AtRAF on hybrid $L_8{}^{A}S_8{}^{t}$ Rubisco a synthetic *Atraf1* gene coding the full length 50.2 kDa *Arabidopsis* RAF1 homolog AY063107 (coding its putative 62 amino acid N-terminal transit peptide sequence; Fig S2B) and a C-terminal 6x histidine tag was cloned 39-bp downstream of At*rbc*L in pLEVAtL. The resulting plasmid, pLEVAtL-R1, was transformed into ^{cm}trL to produce tob^{AtL-R1} lines (Figure 2A). As shown in Fig 1, while most plants only code for one RAF1, tobacco and *Arabidopsis* code two isoforms with the two homologs produced in *Arabidopsis* (~70% identical) only show ~50% identity to the two RAF1 isoforms produced in tobacco (that are 95% identical) (Figure S2C).

In both the tob^{AtL} and tob^{AtL-R1} genotypes the At*rbc*L transgene is regulated by the tobacco *rbc*L promoter, 5'- and 3'-untranslated sequences, and incorporates a downstream promoter-less *aad*A transgene that codes for the spectinomycin resistance used to screen for plastome transformed plantlets (Figure 2A). In tob^{AtL-R1}, the At*raf1* gene is located between both transgenes using an intergenic sequence similar to that used in pLEVL^{Ub}S that produced a bicistronic tobacco *rbc*L-*rbc*S mRNA (23).

Three independent transplastomic tob^{AtL} and tob^{AtL-R1} lines were grown in soil to maturity in air supplemented with 0.5% (v/v) CO_2 and fertilised with wild-type pollen. The increased CO_2 levels were necessary for the survival of the tob^{AtL} lines in soil early during their development as their leaves contained little Rubisco (<3 μmol L-subunits.m^2.s^{-1}), significantly impeding viability and drastically slowing growth in air. In contrast the tob^{AtL-R1} lines grew with greater vigour in air, but still at slow rates.

Comprehensive analyses on the T_1 progeny of the tob^{AtL} and tob$^{AtL\text{-}R1}$ lines were therefore undertaken on plants grown under 0.5% (v/v) CO_2 to ensure their viability.

Variation in the content and catalysis of hybrid $L_8{}^{A}S_8{}^{t}$ Rubisco in the tob^{AtL} and tob$^{AtL\text{-}R1}$ genotypes

RNA blot analyses showed there were large differences in steady state levels of the At*rbc*L mRNAs produced in tob^{AtL} and tob$^{AtL\text{-}R1}$ lines. As observed previously a less abundant At*rbc*L-*aad*A di-cistronic mRNA (~10% that of the At*rbc*L mRNA) was produced in the young tob^{AtL} leaves as a result of inefficient transcription termination by the tobacco *rbc*L 3'UTR (13, 14, 23) (Figure 2B). In contrast, only di-cistronic At*rbc*L-At*raf1* or tri-cistronic At*rbc*L-At*raf1*-*aad*A mRNAs were detected in tob$^{AtL\text{-}R1}$ leaves. Relative to the *rbc*L mRNA levels in the wild type tobacco controls, the total pool of At*rbc*L mRNAs were 25% and 80% lower in the developmentally comparable leaves from tob^{AtL} and tob$^{AtL\text{-}R1}$, respectively (Figure 2B).

In contrast to the scarcity of At*rbc*L transcripts in tob$^{AtL\text{-}R1}$, the levels of hybrid $L_8{}^{A}S_8{}^{t}$ Rubisco (comprising *Arabidopsis* L-subunits and tobacco S-subunits) in the same leaves were >2-fold higher than the $L_8{}^{A}S_8{}^{t}$ content in tob^{AtL} (Figure 2C). This variation in $L_8{}^{A}S_8{}^{t}$ content between each genotype was confirmed by non-denaturing PAGE (ndPAGE). Relative to the level of wild-type L_8S_8 produced in the control, the $L_8{}^{A}S_8{}^{t}$ content in tob^{AtL} and tob$^{AtL\text{-}R1}$ were reduced by ~75% and ~55%, respectively.

Quantifying AtRAF1 production in leaf protein samples was undertaken by immunoblot analysis against varying amounts of purified recombinant AtRAF1 (Figure S3). The AtRAF1 antibody recognised the ~43 kDa AtRAF1 in *Arabidopsis* leaf protein

(Figure 2D), the size expected for mature AtRAF1 after processing of the putative 62 amino acid transit peptide (Figure S1B). The antibody detected nothing in wild-type tobacco consistent with the <50% sequence identity between AtRAF and the two homologs in tobacco (Figure S2C). Compared with *Arabidopsis*, the AtRAF1 produced in $tob^{AtL\text{-}R1}$ leaves was of equivalent size (noting it codes an additional 6x histidines) and produced at similar cellular concentrations (Figure 2D). This indicated the transit peptide processing requirements of AtRAF1 were met by tobacco chloroplast stroma protease(s) and that the levels produced were physiologically comparable to those naturally made in *Arabidopsis*.

The catalytic properties of the hybrid $L_8{}^AS_8{}^t$ were compared with *Arabidopsis* and tobacco Rubisco (Table S1). Significant reductions (24%) in carboxylation rate ($k_C{}^{cat}$) coupled with an improved affinity for CO_2 (*i.e.* a 12% lower K_m for CO_2, K_C) were measured for $L_8{}^AS_8{}^t$ albeit without significant change to its K_m for O_2 (K_O), specificity for CO_2 over O_2 ($S_{C/O}$) or carboxylation efficiency under atmospheric [O_2] ($k_C{}^{cat}/K_C{}^{21\%O2}$).

AtRAF1 forms a stable dimer complex

The AtRAF1 made and purified from *E. coli* could be stably stored at -80°C in buffer containing 20% (v/v) glycerol. Multiple freeze-thaw cycles had no discernible influence on AtRAF1 separation as two bands above the 160 kDa aldolase standard by ndPAGE; a prominent upper band and >90% less abundant lower band (Figure 3A). Immunoblot analysis showed this AtRAF1 oligomer separated at a slower rate than the immune-reactive product detected in *Arabidopsis* leaf protein and the slightly larger His_6-tagged AtRAF1 product (H_6-AtRAF1) produced in $tob^{AtL\text{-}R1}$. The mobility through ndPAGE of

H_6-AtRAF1 from tob$^{AtL\text{-}R1}$ after Ni-NTA affinity purification however matched that of the AtRAF1 purified from *E. coli* (Figure 3A). This suggests the faster migrating, more diffusely separated, AtRAF1 products detected in the *Arabidopsis* and tob$^{AtL\text{-}R1}$ leaf samples might involve complexes with other proteins, the identity of which remain unclarified. In the leaf protein samples, the Rubisco antibody only recognized the L_8S_8 holoenzyme and did not react with any of the products recognized by the RAF1 or CPN antibodies (Figure S3). Similarly, no Rubisco was detected in the protein purified by Ni-NTA from tob$^{AtL\text{-}R1}$ leaves. These finding suggest the AtL-subunits do not form stable interactions with either AtRAF1 or CPN complexes in *Arabidopsis* or tob$^{AtL\text{-}R1}$ leaves.

The migration of proteins through ndPAGE is significantly influenced by their folded quaternary structure which can mislead estimates of molecular size and subunit stoichiometry. For example, the 500 kDa bands for tobacco and *Arabidopsis* Rubisco resolve at different positions following ndPAGE (with the latter resolving at a smaller size to the 440 kDa ferritin protein standard, Figure 3A). We therefore undertook nanoESI-MS analysis of the pure AtRAF1 to accurately determine its subunit stoichiometry. Under non-denaturing conditions, the most abundant ions in the mass spectrum corresponded to a dimer with a molecular mass of approximately 86,871 Da (Figure 3B) consistent with the predicted 43,434 Da for AtRAF1 subunits forming a stable dimer of $(AtRAF1)_2$. This stoichiometry matches that determined for affinity purified RAF1 from *Thermosynechococcus elongatus* cells (22) but contrasts with the trimer structure predicted for RAF1 from maize (19).

Leaf photosynthesis and plant growth are enhanced in tob$^{AtL\text{-}R1}$

Consistent with higher amounts of hybrid $L_8{}^{A}S_8{}^{t}$ made in each tob$^{AtL\text{-}R1}$ line, the leaf photosynthetic CO_2 assimilation rates at varying CO_2 partial pressures (pCO_2) were ~2-fold faster relative to tob^{AtL}, albeit still slower than in wild-type tobacco (Fig 4A). Accordingly, the tob$^{AtL\text{-}R1}$ genotypes grew faster than the tob^{AtL} plants, though again less quickly than the tobacco controls (Fig 4B). Consistent with this faster growth and higher Rubisco contents, the tob$^{AtL\text{-}R1}$ phenotype more closely resembled wild-type with little evidence of the pale green, marginal curling and dimpling leaf phenotype seen for the tob^{AtL} plants. This impaired growth phenotype matches that seen in other tobacco genotypes producing low levels of hybrid Rubisco (*i.e.* <3 µmol sites m^{-2} s^{-1}) comprising tobacco S-subunits and L-subunits from either sunflower (13) or *Flaveria pringlei* (14).

Co-expressing AtRAF1 enhances the post-chaperonin assembly of AtL-subunits into stable $L_8{}^{A}S_8{}^{t}$ complexes

Labelling of intact leaves with ^{35}S-methionine showed varying rates of incorporation into ^{35}S-Rubisco complexes among the different tobacco genotypes (Figure 5A). Compared to tob^{AtL}, the rates of $L_8{}^{A}S_8{}^{t}$ biogenesis were 3-fold faster in the tob$^{AtL\text{-}R1}$, although still 3-fold slower than the rate of L_8S_8 synthesis in the wild-type tobacco controls. Unlabelled methionine 'chase' analyses showed no change in the ^{35}S-Rubisco signal in any tobacco genotype indicating both tobacco L_8S_8 and hybrid $L_8{}^{A}S_8{}^{t}$ complexes were equally stable over the 7 hour analysis period in young upper canopy leaves (Figure 5B).

Discussion

Here we highlight a pivotal role for the chloroplast RAF1 chaperone in Rubisco L-subunit assembly and the underpinning requirement for sequence complementarity between both proteins for optimal rates of L_8S_8 biogenesis. The higher levels and quicker production of $L_8{}^{A}S_8{}^{t}$ Rubisco in $tob^{AtL\text{-}R1}$ leaves (Fig 2C and 5A) and their corresponding faster rates of photosynthesis and growth (Fig 4) relative to the tob^{AtL} genotype underscore the pervasive role that RAF1 plays in the assembly of post-CPN folded L-subunits. This finding advances our understanding of Rubisco biogenesis in leaf chloroplasts and also highlights how chaperone compatibility demands on L-subunit folding and assembly might have constrained Rubisco's catalytic evolution (7, 15).

Our phylogenetic pre-evaluation of parallel evolutionary linkages between the L-subunit and RAF1 and subsequent translational testing of this knowledge by plastome transformation proved highly successful in increasing recombinant Rubisco biogenesis. The specificity shown by Rubisco towards its regulatory protein Rubisco activase (RCA) provides a longstanding example of sequence compatibility requirements between both enzymes (24). Complementarity between residues in the L-subunit N-domain (residues 89 to 94) and those in the specificity H9 helix (resides 317 to 320) of RCA determine the capacity of RCA to stimulate release of inhibitory sugar phosphate molecules from the catalytic sites of Rubisco (25). Similar sequence compliance requirements between L-subunits and other ancillary proteins likely contribute to the low levels of Rubisco from cyanobacteria (12) and other plants (13, 14, 26) that can be produced in tobacco chloroplasts. To what extent expressing the cognate RAF1 proteins for each Rubisco

isoform might augment their biogenesis in tobacco leaves remains untested. Determining the extent of parallel evolutionary linkages between the L-subunit and other molecular partners considered influential to Rubsico biogenesis (eg. CPN, BSDII, RBCX, RAF2) may help identify those whose co-expression might augment recombinant Rubisco assembly in chloroplasts and other expression systems. This approach is particularly pertinent to the ongoing efforts to design and express more efficient Rubisco variants in crop plants (6).

Our analysis of AtRAF1 produced in *E. coli* indicates that it forms a stable dimer that differs in its migration size through ndPAGE to the RAF1 in soluble leaf cellular protein extract (Fig 3A). This suggests RAF1 in chloroplasts might interact with other proteins or cofactors that alter quaternary structure or/and prevent dimer formation due to assembly with other proteins that are sufficiently stable to ndPAGE separation, but not to Ni-NTA purification where $(RAF1)_2$ oligomers matching those purified from *E. coli* are formed. Recent analysis of formaldehyde-treated maize leaf protein indicated RAF1 may interact with RAF2 and BSDII (20). Whether such interactions are responsible for the different migration rates through ndPAGE is a possibility that remains to be tested. Resolving the crystal structure for the $(RAF1)_2$ complex should help reveal its potential for forming alternative quaternary structures that might explain its alternative ndPAGE separation patterns and propensity to separate as an apparently larger sized complex that has previously been interpreted as a trimer (19, 20). For example, are the variations in $(RAF1)_2$ separation by ndPAGE due to its capacity to form "closed" and "open" conformations or/and from interactions with ancillary proteins or co-factors?

Constraints on the steady state At*rbc*L mRNA levels in tob$^{AtL\text{-}R1}$ leaves appear a leading cause to limiting $L_8{}^{A}S_8{}^{t}$ biogenesis. The steady state pool of At*rbc*L mRNA in tob$^{AtL\text{-}R1}$ leaves was reduced 5-fold relative to the tobacco *rbc*L mRNA levels (Fig. 2B), but still managed to produce $L_8{}^{A}S_8{}^{t}$ at half the levels of L_8S_8 made in wild-type (Fig 2C). This would suggest producing more hybrid $L_8{}^{A}S_8{}^{t}$, possibly matching wild-type Rubisco levels, would be feasible by enhancing At*rbc*L mRNA levels. The operon structure in tob$^{AtL\text{-}R1}$ matches that used previously in the transplastomic LEVUbS tobacco genotype. As seen in tob$^{AtL\text{-}R1}$ leaves (Fig 2B), the LEVUbS leaves also produced a di-cistronic *rbc*L-Ub*rbc*S mRNA and a 5- to 6-fold less abundant tri-cistronic *rbc*L-Ub*rbc*S-*aad*A transcript; however they were produced at levels that matched the *rbc*L mRNA content in wild-type (23). This suggests the At*raf1* transgene likely destabilizes the di- and tri-cistronic At*rbc*L transcripts produced in tob$^{AtL\text{-}R1}$. Future RAF1 transplastomic studies should therefore consider equipping the *raf1* transgene with separate promoter/terminator regulatory elements to those controlling *rbc*L expression. Alternatively a small RNA intercistronic expression element (IEE) between the *rbc*L and *raf1* transgenes that has been shown to trigger processing of polycistronic transcripts into more stable and translatable smaller transcripts could be included (27).

Previous studies of hybrid Rubiscos comprising plant L-subunits have shown the pervasive role of the L-subunit on shaping catalysis (13, 14, 28). Here a modest, yet significant, reduction in $k_C{}^{cat}$ and improvement in K_C was found for the $L_8{}^{A}S_8{}^{t}$ Rubisco relative to the native *Arabidopsis* and tobacco enzymes, which have comparable catalytic constants at 25°C (Table S1). This catalytic variability of $L_8{}^{A}S_8{}^{t}$ Rubisco likely arises

from complementarity differences between *Arabidposis* and tobacco S-subunits, consistent with a growing appreciation of the influential role the S-subunits can have on catalysis (6, 29).

Here we demonstrate the importance of a chaperone compatibility to enhancing recombinant Rubisco production in tobacco plastids. The finding enhances the potential for bioengineering Rubisco in chloroplasts and provides mechanistic evidence for the role of RAF1 in L-subunit assembly. Future applications of this co-engineering approach will focus on identifying ways to more efficiently co-express Rubisco L-subunits and their complementary RAF1(s) without compromising leaf *rbc*L mRNA pools. Extending this transplastomic co-expression method to other Rubisco chaperones – BSDII, RBCX, and RAF2 – may prove a useful approach for determining their biochemical function in chloroplasts.

Materials and Methods

Bioinformatics Analyses

Full length *raf1* and *rbc*L sequences from 26 plant, three algal and 46 cyanobacterial genomes were obtained from NCBI (http://www.ncbi.nlm.nih.gov) and Phytozome (http://www.phytozome.net) using the BLAST algorithm (Table S2). Phylogenetic trees of the translated proteins were constructed by the RAxML program (30) using the Maximum Likelihood method with the following parameters: the Dayhoff model with gamma distributed rates, partial deletion, and bootstrap (1000 replicates; random seed). L-subunit and RAF1 phylogenetic trees were compared using the Mirrortree server (31). Pairwise non-synonymous (leading to amino acid substitutions) and synonymous (selectively neutral) sequence distances were calculated using the PAML package (32). We used the Mantel test to compute the Pearson correlation coefficient R. The chloroplast gene, *mat*K, encoding maturase K (absent in most cyanobacteria genomes) which doesn't interact with Rubisco, was included as a negative control.

Tobacco plastome transformation and growth

The *rbc*L gene from *Arabidopsis* was PCR amplified from leaf genomic DNA with primers 5'NheIrbcL (14) and 3'AtSalIrbcL (5'-TGTCGACTGTTTTTATCTCTTCTTATCCTTATCCT-3') and the 1439-bp *Nhe*I-*Sal*I At*rbc*L product cloned into pLEV4 (14) to give pLEVAtL (genbank KP635965). A synthetic At*raf1* gene whose codon use matched tobacco *rbc*L was synthesised by GenScript and cloned downstream of At*rbc*L in pLEVAtL using the intergenic sequence used in pLEVLUbS (23) to give pLEVAtL-R1 (genbank KP635964). pLEVAtL and

pLEVAtL-R1 were each biolistically transformed into five leaves of the tobacco-masterline ^{cm}trL as described in (23) with 4 and 7 spectinomycin-resistant plants, respectively, obtained. Three independent plastome transformed lines of each genotype were grown to maturity in soil in a growth atmosphere supplemented with 0.5% (v/v) CO_2 as described (13) and fertilised with wild-type pollen. The resulting T_1 progeny were used for all analyses.

RNA blot, PCR, protein and PAGE analyses

Total leaf genomic DNA was isolated using the DNeasy® Plant Mini Kit and used to PCR amplify and sequence the transformed plastome region using primers LSH and LSE (14) (Fig 1B). Total RNA extracted from 0.5 cm^2 leaf discs was separated on denaturing formaldehyde gels, blotted onto Hybond-N nitrocellulose membrane (GE healthcare) and probed with the ^{32}P-labelled 5'UTR probe (Fig 2A) as described (13). The preparation, quantification (against BSA) of soluble leaf protein and analysis by SDS-PAGE, ndPAGE and immunoblot analysis was performed as described (33).

Rubisco content and catalysis

Rates of Rubisco fixation in soluble protein extracts from 3 different leaves of each tobacco genotype and *Arabidopsis* were measured under varying concentrations of $NaH^{14}CO_3$ (0 to 43 µM) and O_2 (0 to 25% (v/v)) and the Michaelis constants (K_m) for CO_2 (K_C) and O_2 (K_O) determined from the fitted data (14). The maximal rate of carboxylation (V_C) was extrapolated from the Michaelis-Menten fit and then divided by the amount of Rubisco active sites quantified by [^{14}C]-2-CABP binding (33, 34) to

determine the turnover rate (k^{C}_{cat}). Rubisco CO_2/O_2 specificity ($S_{C/O}$) was measured using ion exchange purified protein as described (13).

Growth and photosynthesis analysis

All plants were grown in a growth chamber at 25°C in air containing 0.5% (v/v) CO_2 as described (13). Leaf photosynthesis rates were measured using a LI-6400 gas-exchange system (LI-COR) on the 5th upper canopy leaf of each tobacco genotype once they had reached comparable stages of physiological development.

Recombinant RAF1 and CPN60α purification and antibody production

Genes coding *Arabidopsis* RAF1 (AY063107) and Chaperonin 60α2 (NM_121887) were cloned into plasmid pHueAct and expressed as N-terminal 6-Histidine-ubiquitin (H_6Ub) tagged proteins in BL21(DE3) cells and purified by affinity chromatography (Figure S2). Antibodies to both purified proteins were raised in rabbits.

Mass spectrometry

Purified AtRAF1 stored at -80°C in buffer containing 20% (v/v) glycerol was dialysed (14000 MWCO) against 100 mM ammonium acetate buffer adjusted to pH 7.2. The protein concentration was measured using a Nanodrop2000c (Thermo Fisher Scientific) and adjusted to 3 µM (monomer concentration) prior to mass spectrometry. Positive ion nanoESI mass spectra were acquired using a Waters (Manchester, UK) Synapt^{TM} HDMS^{TM} fitted with a Z-spray nanoESI source. Spectra were acquired using a MCP potential of 1850 V, capillary voltage of 1.5 kV, extraction cone voltage of 4 V and sampling cone voltages of 30, 80 and 150 V. The source temperature was set to 30 °C,

the nanoflow back pressure to 0.1 bar and the backing pressure to 3.93 mbar. The trap and transfer collision energies were 6.0 V and 4.0 V, respectively. Spectra were acquired over the 500 - 10000 *m/z* range and 40-50 acquisitions. The instrument was calibrated using a CsI solution (10 mg/mL in water).

Pulse-chase labelling with ^{35}S

Plants of comparable size (~38 cm in height) stored overnight in a darkened laboratory were equilibrated for 15 min with ~500 μmol photons m^2 s^{-1} illumination (at the surface of the youngest near fully expanded leaf sampled). Upper canopy leaves of equivalent age were infiltrated through the abaxial stomata by syringe (see Fig. S5) with 3 to 4 mL of Trans^{35}S-label (ICN) diluted to 0.25 mCi ml^{-1} (9.25 MBq ml^{-1}) with infiltration buffer (10 mM MES-NaOH pH 5.5, 10 mM $MgSO_4$). This process took 45 to 60 sec. Leaf discs (0.5 cm^2) were collected after 15, 30 and 45 min and frozen in liquid nitrogen. After 60 min the leaves were infiltrated with infiltration buffer containing 10 mM methionine and leaf samples taken after 2, 4 and 7 h. The soluble leaf protein was separated by ndPAGE, the proteins fixed by Coomassie staining before drying the gels and exposing to a Storage Phosphor screen GP (Kodak) for 2 days. The autoradiograph signals were visualized using a PharosFX Molecular Imager and quantified with Quantity One software (Biorad).

Affinity purification of 6xHis-tagged AtRAF1 from tob$^{AtL\text{-}R1}$ leaves

Soluble leaf protein from tob$^{AtL\text{-}R1}$ and wild-type tobacco (negative control) was purified by Ni^{2+}-nitrilotriacetic acid (Ni-NTA) agarose (Qiagen) chromatography and analysed by

SDS PAGE, ndPAGE and immunoblotting for evidence of stable interactions between AtRAF, AtL-subunits and CPN (Fig S4).

Acknowledgements

This research was supported by Australian Research Council grants FT0991407 (SW), CE140100015 (SW), LE0882289 (JB) and the Bill and Melinda Gates Foundation funded project Realizing Increased Photosynthetic Efficiency (RB, MK).

References

1. Edgerton MD (2009) Increasing Crop Productivity to Meet Global Needs for Feed, Food, and Fuel. *Plant Physiol* 149(1):7-13.
2. Ziska LH, *et al.* (2012) Food security and climate change: on the potential to adapt global crop production by active selection to rising atmospheric carbon dioxide. *Proceedings of the Royal Society B: Biological Sciences* 279(1745):4097-4105.
3. Evans JR (2013) Improving photosynthesis. *Plant Physiol* 162(4):1780-1793.
4. Peterhansel C & Offermann S (2012) Re-engineering of carbon fixation in plants - challenges for plant biotechnology to improve yields in a high-CO_2 world. *Curr Opin Biotech* 23(2):204-208.
5. Zhu X-G, Long SP, & Ort DR (2010) Improving photosynthetic efficiency for greater yield. *Ann Rev Plant Biol* 61(1):235-261.
6. Parry MAJ, *et al.* (2013) Rubisco activity and regulation as targets for crop improvement. *J Exp Botany* 64(3):717-730.
7. Whitney SM, Houtz RL, & Alonso H (2011a) Advancing our understanding and capacity to engineer nature's CO_2-sequestering enzyme, Rubisco. *Plant Physiol* 155(1):27-35.
8. Andersson I & Backlund A (2008) Structure and function of Rubisco. *Plant Physiol Biochem* 46(3):275-291.
9. Evans J & Seemann J (1989) The allocation of nitrogen in the photosynthetic apparatus: costs, consequences and control. *Photosynthesis*, ed Briggs WR (Alan R Liss, Inc, New York), pp 183-205.
10. Zhu XG, Portis AR, & Long SP (2004) Would transformation of C_3 crop plants with foreign Rubisco increase productivity? A computational analysis extrapolating from kinetic properties to canopy photosynthesis. *Plant Cell Environ* 27(2):155-165.
11. Nishimura K, Ogawa T, Ashida H, & Yokota A (2008) Molecular mechanisms of Rubisco biosynthesis in higher plants. *Plant Biotech.* 25:285-290.
12. Lin MT, Occhialini A, Andralojc PJ, Parry MAJ, & Hanson MR (2014) A faster Rubisco with potential to increase photosynthesis in crops. *Nature* 513(7519):547-550.
13. Sharwood R, von Caemmerer S, Maliga P, & Whitney S (2008) The catalytic properties of hybrid Rubisco comprising tobacco small and sunflower large subunits mirror the kinetically equivalent source Rubiscos and can support tobacco growth. *Plant Physiol* 146:83-96.

14. Whitney SM, *et al.* (2011b) Isoleucine 309 acts as a C_4 catalytic switch that increases ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) carboxylation rate in *Flaveria*. *Proc Nat Acad Sci* 108(35):14688-14693.

15. Mueller-Cajar O & Whitney SM (2008) Directing the evolution of Rubisco and Rubisco activase: first impressions of a new tool for photosynthesis research. *Photosynth Res* 98(1-3):667-675.

16. Hemmingsen SM, *et al.* (1988) Homologous plant and bacterial proteins chaperone oligomeric protein assembly. 333:330-334.

17. Tsai Y-CC, Mueller-Cajar O, Saschenbrecker S, Hartl FU, & Hayer-Hartl M (2012) Chaperonin cofactors, Cpn10 and Cpn20, of green algae and plants function as hetero-oligomeric ring complexes. *J Biol Chem* 287(24):20471-20481.

18. Bracher A, Starling-Windhof A, Hartl FU, & Hayer-Hartl M (2011) Crystal structure of a chaperone-bound assembly intermediate of form I Rubisco. *Nat Struct Mol Biol* 18(8):875-880.

19. Feiz L, *et al.* (2012) Ribulose-1,5-Bis-Phosphate Carboxylase/Oxygenase Accumulation Factor1 Is Required for Holoenzyme Assembly in Maize. *Plant Cell* 24(8):3435-3446.

20. Feiz L, *et al.* (2014) A protein with an inactive pterin-4a-carbinolamine dehydratase domain is required for Rubisco biogenesis in plants. *Plant J* 80(5):862-869.

21. Brutnell TP, Sawers RJH, Mant A, & Langdale JA (1999) Bundle sheath defective2, a novel protein required for post-translational regulation of the *rbc*L gene of maize. *Plant Cell* 11(5):849-864.

22. Kolesinski P, Belusiak I, Czarnocki-Cieciura M, & Szczepaniak A (2014) Rubisco Accumulation Factor 1 from *Thermosynechococcus elongatus* participates in the final stages of ribulose-1,5-bisphosphate carboxylase/oxygenase assembly in *Escherichia coli* cells and in vitro. *FEBS J* 281(17):3920-3932.

23. Whitney SM & Sharwood RE (2008) Construction of a tobacco master line to improve Rubisco engineering in chloroplasts. *J Exp Bot* 59(7):1909-1921.

24. Mueller-Cajar O, Stotz M, & Bracher A (2014) Maintaining photosynthetic CO_2 fixation via protein remodelling: the Rubisco activases. *Photosynth Res* 119(1-2):191-201.

25. Stotz M, *et al.* (2011) Structure of green-type Rubisco activase from tobacco. *Nature Struct Mol Biol* 18(12):1366-1370.

26. Whitney SM, Baldet P, Hudson GS, & Andrews TJ (2001) Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *Plant J* 26(5):535-547.

27. Lu Y, Rijzaani H, Karcher D, Ruf S, & Bock R (2013) Efficient metabolic pathway engineering in transgenic tobacco and tomato plastids with synthetic multigene operons. *Proc Nat Acad Sci* 110(8):623–632.

28. Zhang XH, *et al.* (2011) Hybrid Rubisco of tomato large subunits and tobacco small subunits is functional in tobacco plants. *Plant Sci* 180(3):480-488.

29. Ishikawa C, Hatanaka T, Misoo S, Miyake C, & Fukayama H (2011) Functional incorporation of sorghum small subunit increases the catalytic turnover rate of Rubisco in transgenic rice *Plant Physiol* 156:1603-1611.

30. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312-1313.

31. Ochoa D & Pazos F (2010) Studying the co-evolution of protein families with the Mirrortree web server. *Bioinformatics* 26(10):1370-1371.

32. Xu B & Yang Z (2013) pamlX: A graphical user interface for PAML. *Mol Biol Evol* 30(12):2723-2724.

33. Whitney SM & Sharwood RE (2007) Linked Rubisco subunits can assemble into functional oligomers without impeding catalytic performance. *J Biol Chem* 282(6):3809-3818.

34. Whitney SM & Andrews TJ (2001) Plastome-encoded bacterial ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) supports photosynthesis and growth in tobacco. *Proc Natl Acad Sci* 98(25):14738-14743.

Figure Legends

Figure 1. RAF1 and Rubisco L-subunits phylogenies of plants, green algae and β-cyanobacteria. (A) Condensed RAF1 and L-subunit (RBCL) ML trees assembled using RAxML v.8. Full ML trees are shown in Fig S1 and sequence accessions listed in Table S2. (B) Correlations of pairwise non-synonymous d_N (leading to amino acid substitutions) and synonymous d_S (selectively neutral) distances for RAF1, L-subunit and maturase K (*mat*K, an un-associated chloroplast made protein; negative control) across green plants and algae (all significant at $p < 0.0001$).

Figure 2. Transplastomic tobacco generation and analysis of Rubisco and AtRAF1 expression. (A) The transforming plasmids pLEVAtL (genbank KP635965) and pLEVAtL-R1 (genbank KP635964) contain homologous plastome flanking sequence (indicated by dashed lines, numbering indicates region of sequence integration relative to *N. tabacum* (wt) plastome sequence; genbank Z00444) that directed integration of the At*rbc*L or At*rbc*L-*raf*1 transgenes and a promoter-less *aad*A selectable marker gene into the ^{cm}trL tobacco genotype plastome (23) to produce lines tob^{AtL} and tob$^{AtL\text{-}R1}$. The tobacco *rbc*L promoter/5'UTR (P) and first 42 nucleotides of wt *rbc*L sequence are conserved in each tobacco genotype. This sequence corresponds to the 5UTR probe (14) with the expected mRNA species identified by the probe shown (dashed arrows). t, *rps16* 3'UTR, *T, psb*A 3'UTR, T, *rbc*L 3'UTR. (B) Detection of the various *rbc*L coding mRNA transcripts by the 5UTR probe in total RNA from 6 mm^2 of young, nearly fully expanded leaves (14 to 16 cm in diameter) from comparable positions in the canopy of 32 ± 4 cm tall plants of independent T_1 transformed lines and 3 wt controls. (C) Variation in

the mean (±SD) Rubisco content in tobacco leaves analysed in (B) and those from three *Arabidopsis* (*At*) leaves as quantified by ^{14}C-CABP binding. Shown is an example ndPAGE analysis of the leaf protein used to confirm the varied levels of L_8S_8 Rubisco. (D) AtRAF1 production in the *At*, wt and tob$^{AtL\text{-}R1}$ leaf protein analysed in (C) was quantified by SDS PAGE immunoblot analysis (example shown) against known amounts of purified AtRAF1 (Fig S2B). *, the AtRAF1antibody does not recognise tobacco RAF1.

Figure 3. AtRAF1 stably assembles as a dimer. (A) ndPAGE analyses reproducibly showed recombinant AtRAF1 oligomers purified from *E. coli* (pure, Fig S2A) was highly stable and separated at the same position above aldolase (160 kDa) in the marker protein standards (m) as Ni^{2+}-nitrilotriacetic acid agarose (Ni-NTA) agarose purified His_6-tagged AtRAF1 complexes (AtRAF1^{H6}) from tob$^{AtL\text{-}R1}$ (t$^{AtL\text{-}R1}$) leaves (see Fig. S4 for detail). In *Arabidopsis* (*At*) and t$^{AtL\text{-}R1}$ leaf soluble protein the AtRAF1 and larger AtRAF1^{H6} separated as smaller, more diffuse protein complexes of unknown content (indicated by *). Variations in the amount of sample loaded per lane relative to the Coomassie stained gel are shown in parentheses. (B) NanoESI mass spectrum of pure AtRAF1 (3.2 μM; buffer exchanged into 0.1 M ammonium acetate, pH 7.2; cone voltage, 80 V) shows that the most abundant isoform was the dimer (*i.e.* $(AtRAF)_2$), with ions of low abundance from the monomer, and small amounts of unfolded monomer and dimer. The folded dimer was the most abundant isoform under cone voltages of 30 – 150 V). ● Folded dimer $(AtRAF)_2$, ♦ Folded monomer AtRAF, ○ Unfolded dimer $(AtRAF)_2$, ◊ Unfolded monomer AtRAF.

Figure 4. AtRAF1 improved leaf photosynthesis and growth in tob$^{AtL\text{-}R1}$. (A) Leaf gas exchange measurements of CO_2-assimilation rates at 25°C under varying intercellular CO_2 pressures (C_i) made at 1000 µmol quanta m^{-2} s^{-1} illumination. Shown are the average of 3 measurements (±SD) made on the leaves analysed in Fig 2. (B) Comparison of the faster growth (as a function of plant height ±SD) of the tob$^{AtL\text{-}R1}$ lines (n=3) relative to tob^{AtL} (n=3) at 25°C in a growth cabinet in air with 0.5% (v/v) CO_2 under ~400 ± 100 µmol quanta m^{-2} s^{-1} illumination. Both transplastomic genotypes grew slower than wild-type tobacco (wt, n=3). (C) Phenotype of the plants at the respective age post-cotyledon emergence (pce).

Figure 5. AtRAF1 stimulated assembly of Rubisco. ^{35}S-Met 'pulse' - unlabelled-Met 'chase' analysis of hybrid $L_8{}^{A}S_8{}^{t}$ Rubisco synthesis and turnover relative to tobacco L_8S_8 Rubisco performed on young attached leaves under constant illumination (~500 µmol quanta m^{-2} s^{-1}, see Figure S5). (A) Autoradiography signals of ndPAGE separated soluble protein from 6 mm^2 of leaf taken 15, 30 and 45 min after infiltration with ^{35}S-methionine showing increasing ^{35}S incorporation into L_8S_8 Rubisco. Plotted are the average densitometry signals for L_8S_8 Rubisco at each time point (n = 3 ± SD) relative to the average of the 45 min wt sample signals. Rates of L_8S_8 synthesis extrapolated from linear fits to the normalised data were 27 x 10^{-4} (r^2 = 0.999, 78 x 10^{-4} (r^2 = 0.997) and 229 x 10^{-4} (r^2 = 1.000) for the tob^{AtL} (●), tob$^{AtL\text{-}R1}$ (○) and wt (□) leaves respectively. (B) ndPAGE analyses made on soluble protein from the same leaves taken 2h, 4h and 7h after a 'chase' infiltration with 10 mM unlabelled-methionine. No discernible changes in the

densitometry of either hybrid $L_8^{A}S_8^{t}$ or wild type L_8S_8 Rubisco autoradiography signals were detected indicative of little, or no, Rubisco turnover during this period.

Figure 1

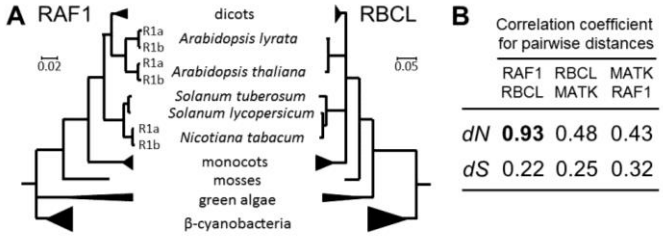


Figure 2

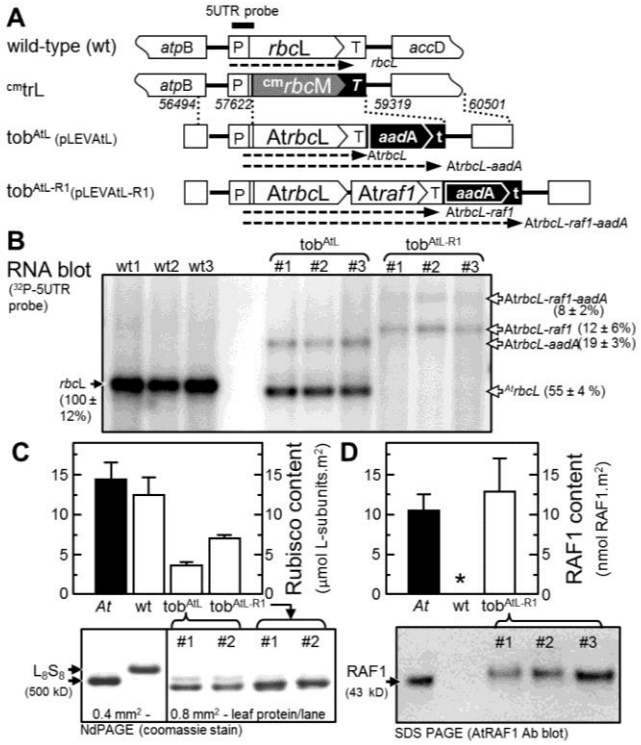


Figure 3

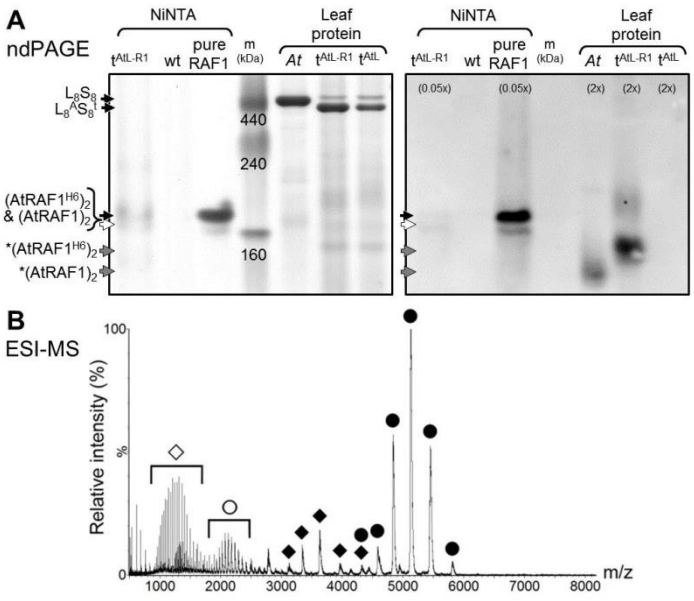


Figure 4

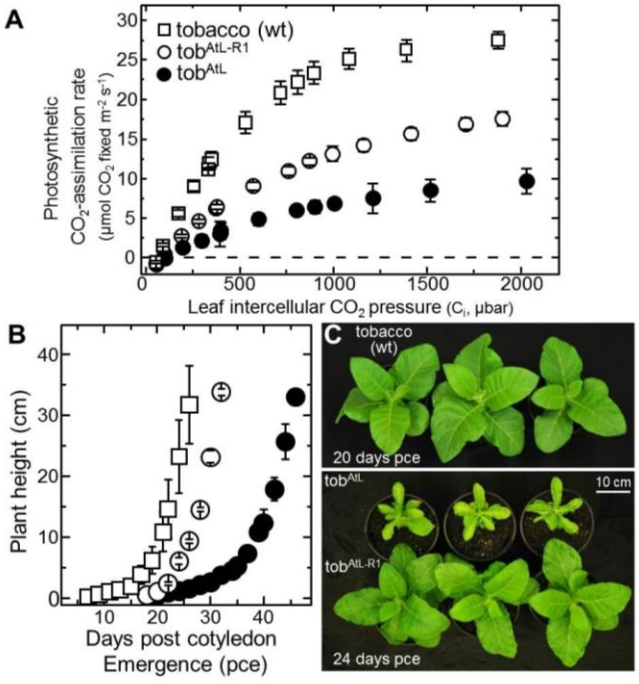
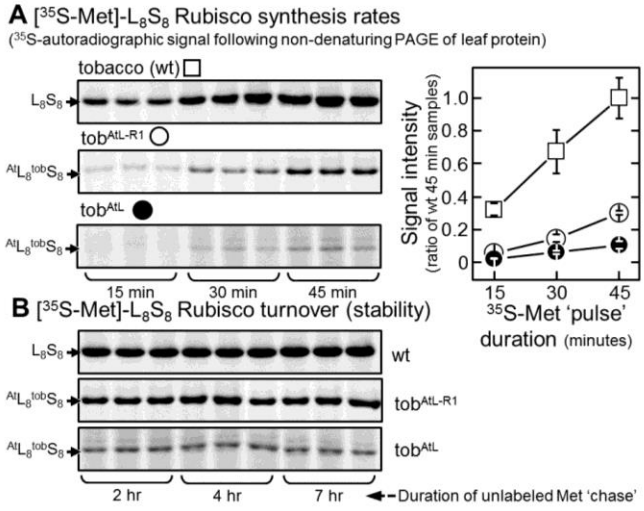


Figure 5


A [35S-Met]-L8S8 Rubisco synthesis rates
(35S-autoradiographic signal following non-denaturing PAGE of leaf protein)
tobacco (wt)
L8S8
tobAtL-R1
AtL8tobS8
tobAtL
AtL8tobS8
15 min
30 min
45 min
Signal intensity
(ratio of wt 45 min samples)
1.0
0.8
0.6
0.4
0.2
0
15
30
45
35S-Met 'pulse'
duration (minutes)
B [35S-Met]-L8S8 Rubisco turnover (stability)
L8S8
wt
AtL8tobS8
tobAtL-R1
AtL8tobS8
tobAtL
2 hr
4 hr
7 hr
Duration of unlabeled Met 'chase'

Classification: Biological Sciences –Plant Biology

Improving recombinant Rubisco biogenesis, plant photosynthesis and growth by co-expressing its ancillary RAF1 chaperone

RAF1 co-expression enhances foreign Rubisco assembly

Spencer M. Whitney*‡, Rosemary Birch*, Celine Kelso†, Jennifer L. Beck†, Maxim V. Kapralov*

* Research School of Biology, The Australian National University, Acton, Australian Capital Territory 2601, Australia. †School of Chemistry, University of Wollongong, New South Wales 2522, Australia

‡ Corresponding author; Research School of Biology, Australian National University, Acton, Australian Capital Territory 2601, Australia; Tel: +61-2-6125-5073; E-mail: spencer.whitney@anu.edu.au

Number of text pages: 6

Number of Supplemental figures: 5

Number of Supplemental Tables: 2

Supplemental data

Figure S1. RAF1 and Rubisco L-subunits phylogenies of plants, green algae and β-cyanobacteria.

(A) ML trees assembled under the Dayhoff model implemented in RAxML v.8 (1) using translated amino acid sequences from the full length *raf1* and *rbc*L genes listed in Table S2. Posterior probability (PP) values are shown above tree branches; all clades with PP < 0.5 have been dissolved.

Figure S2. Sequence comparison of the Rubisco L-subunit and RAF1 isoforms in tobacco and *Arabidopsis*.

Alignment of (A) Rubisco L-subunits and (B) RAF1 homologs from *Arabidopsis thaliana* and *Nicotiana tabacum*. Tobacco *rbc*L (NC_001879) and *Arabidopsis rbc*L and *raf1* (ArthCp030, AT3G04550, AT5G28500) sequences were obtained from GenBank. The tobacco RAF1 sequences (*Nt-R1a* and *Nt-R1b*) were derived from the assembly of Illumina RNA-Seq transcriptome data of *N. tabacum* cv. K326 (Sequence Read Archive accession code SRP029184; (2)) using CLC Genomics Workbench 7.0.3 (http://www.clcbio.com) software. (C) Sequence identities of the different RAF1 homologs after Clustal W alignment both with and without (shade grey) their predicted transit peptide coding sequences (highlighted red in panel B).

Figure S3- CPN60α and ^{At}RAF1 purification and quantification by immunoblot analysis.

The mature coding sequence CPN60α1 (Genbank NP_197383.1, At5g18820) from *Arabidopsis* (*i.e.* spanning amino acids 36 to 578 to exclude part or all of the chloroplast targeting sequence) was amplified by RT-PCR (SuperScript III Reverse Transcriptase, Life Technologies) using leaf RNA extracted using TRIzol Reagent (Life Technologies) and primers 5'SacIIAtCPN60α (5'-CCGCGGTGGAATGGGAGCTAAGAGAATACTATAC-3') and 3'HindIII AtCPN60α (5'-AAGCTTATGATGTGGGTATGCCAGG-3'). The amplified 1637-bp *Sac*II-*Hind*III product was cloned in frame with the N-terminal 6x-histidine (H_6)-Ub fusion peptide in plasmid pHue (3) to give plasmid pHueCPN60α. Similarly, the synthetic ^{At}*raf1* gene in pLEVAtL-RAF1 (Figure 1A) was amplified with primers 5'SacIIAtRAF1 (5'-CCGCGGTGGAATGGCTCCTCTTAAATCTTTGATT-3') and 3'HindIIIAtRAF1 (5'-AAGCTTCTCGAGATCCCAATTTTGATG-3') and the 1364-bp *Sac*II-*Hind*III fragment cloned into pHue to give pHueAtRAF1. *Escherichia coli* BL21 (DE3) cells transformed with plasmids pHueAtRAF1 and pHueCPN60α were grown at 28°C on a rotary shaker (150 rpm) in 0.5 L of Luria-Bertani medium containing 200 μg/mL ampicillin. At an A_{600} of 1.0 isopropyl-b-$_D$-thiogalactopyranoside was added to 0.5 mM. After 6h, the cells were harvested by centrifugation (3,300 *g,* 10 min, 4°C) and resuspended in 10 mL of ice-cold extraction buffer (0.1 M Tris-HCl, pH 8.0, 0.3 M NaCl, 1 mM PMSF, 5 mM mercaptoethanol) and lysed by passage through a pre-chilled French pressure cell at 140 MPa. The extract was centrifuged (33,000 *g,* 10 min, 4°C) and the (H_6)-Ub-RAF1 and (H_6)-UbCPN60α proteins purified by Ni^{2+}-nitrilotriacetic acid (Ni-NTA) agarose

(Qiagen) chromatography, eluted in imidazole buffer (extraction buffer with 0.2M imidazole) and the (H_6)-Ub sequences removed with a (H_6)-Ub-protease as described [Baker REF] before dialysing into storage buffer (40 mM EPPS-NaOH, pH8, 8 mM MgCl2, 0.8 mM EDTA, 20% (v/v) glycerol) and storing at -80°C.

(A) Protein samples during the purification were diluted with 0.25-volumes 4x SDS reducing buffer and analysed by SDS PAGE as described (4). (B) The ^{At}RAF1 content in soluble protein from known leafs areas were calculated by immuno-blot densitometry analysis against known amounts of purified ^{At}RAF1 (quantified against BSA standards) separated in parallel by SDS PAGE.

Figure S4. PAGE analysis of NiNTA purified and total soluble leaf protein from *Arabidopsis* and the different tobacco genotypes.

(A) ndPAGE and (B) SDS PAGE analysis of soluble leaf protein (from Arabidopsis (*At*), tob$^{AtL\text{-}R1}$ and tob^{AtL}) and Ni^{2+}-nitrilotriacetic acid agarose (Ni-NTA) purified protein from *E. coli*-pHueAtRAF1 cells (Fig S2), tobacco (wt) and tob$^{AtL\text{-}R1}$ leaves. Variations in the amount of sample loaded per lane relative to the Coomassie stained gel are shown in parentheses. For NiNTA purification ~2g of tob$^{AtL\text{-}R1}$ and wild-type tobacco leaves were homogenised in 20 mL extraction buffer (0.1 M Tris-HCl, pH 8.0, 0.3 M NaCl, 5% v/v glycerol, 1% w/v PVPP, 1 mM PMSF, 5 mM mercaptoethanol) using 40 mL Wheaton glass homogenisers then centrifuged (16,500 *g*, 10 min, 2°C). The soluble protein was transferred to a 10 mL Econo column (Promega) containing a 1 mL bed volume of Ni-NTA agarose (Qiagen). After the sample had passed through the resin it was washed with

20 bed volumes of extraction buffer (no PVPP or mercaptoethanol). The bound protein was collected in 0.8 mL of elution buffer (0.1 M Tris-HCl, pH 8.0, 0.3 M NaCl, and 200 mM imidazole) and the proteins separated by PAGE as described (4). Immunoblot analysis confirmed the ^{At}RAF1 purified from tob$^{AtL\text{-}R1}$ comprised two similar sized bands that matched the size of those purified from *E. coli*. In the *At* and tob$^{AtL\text{-}R1}$ soluble leaf protein samples the native ^{At}RAF1 and slightly larger recombinant ^{At}RAF1^{H6} products are seen as more diffuse bands of lower apparent molecular size. No Rubisco or CPN60α subunits were detected in the NiNTA purified protein from tob$^{AtL\text{-}R1}$ or wild-type. Only the ^{At}RAF1 protein was visually unique in the Coomassie stained NiNTA purified protein from tob$^{AtL\text{-}R1}$ suggesting it does not stably interact with any other tobacco chloroplast protein to any significant extent, although this requires closer proteomic scrutiny.

Figure S5. ^{35}S-labeling of Rubisco in attached tobacco leaves by a direct infiltration approach.

Due to significant variations in Rubisco expression down the canopy of tobacco (5), significant care was taken to perform the ^{35}S-infiltration experiments on leaves of comparable developmental status and positioning in the upper canopy. (A) The plants analysed were all of comparable size with infiltration experiments performed on the youngest near fully expanded leaf (the fifth from the top of the canopy, indicated by white arrow) where the intercellular air spaces are optimally developed for fast and efficient liquid infiltration. (B) Showing the regions of the leaves towards the tip that

were infiltrated in the experiment and the sampling protocol undertaken during both the ^{35}S-methionine labeling ('pulse') and ensuing 10 mM methionine 'chase' period.

References

1. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312-1313.
2. Sierro N, *et al.* (2014) The tobacco genome sequence and its comparison with those of tomato and potato. *Nat Commun* 5.
3. Baker RT, *et al.* (2005) Using Deubiquitylating Enzymes as Research Tools. *Methods in Enzymology*, ed Raymond JD (Academic Press), Vol Volume 398, pp 540-554.
4. Whitney SM & Sharwood RE (2007) Linked Rubisco subunits can assemble into functional oligomers without impeding catalytic performance. *J Biol Chem* 282(6):3809-3818.
5. Pengelly JJ, *et al.* (2014) Transplastomic integration of a cyanobacterial bicarbonate transporter into tobacco chloroplasts. *J Exp Bot* 65(12):3071-3080.

Table S1. Rubisco catalysis comparison

Plant source	tobacco	*Arabidopsis*	tob$^{AtL\text{-}R1}$
k_C^{cat} (s^{-1})	3.1 ± 0.1	3.0 ± 0.2	2.3 ± 0.3*
K_C (μM)	9.7 ± 0.2	9.8 ± 0.3	8.6 ± 0.2*
K_O (μM)	174 ± 16	192 ± 17	221 ± 16
$k_C^{cat}/K_C^{21\%O2}$ ($mM^{-1}{\cdot}s^{-1}$)	138	125	126
$S_{C/O}$ ($mol.mol^{-1}$)	82 ± 1	80 ± 2	80 ± 3

*Significance variation ($p<0.05$) determined by T-test. $K_C^{21\%O2}$, the apparent K_m for CO_2 (K_C) at atmospheric [O_2] (assumed 252 μM at 25°C) calculated as $K_C(1+[O_2]/K_O)$.

Table S2. List of species and accession numbers for the *raf1* and *rbc*L sequences from 26 plant, 3 algal and 46 cyanobacteria genomes used to construct the ML trees in Fig. S1. Two gene copies of *raf1* were found in five plant species (including tobacco and *Arabidopsis*, see Fig S2B), and one copy in all other species. Accession numbers are also shown for the chloroplast *mat*K sequences that were used as a negative control when testing for putative *raf1* and *rbc*L co-evolution by correlating their pairwise non-synonymous (leading to amino acid substitutions) and synonymous (selectively neutral) distances across green plants and algae (see Fig 1B).

Organism	***raf1***	***rbcL***	***matK***
Angiosperms			
Arabidopsis lyrata	XM_002882316; XM_002872267	XM_002888303	AF144342
Arabidopsis thaliana	BT015787; AY063107	U91966ATU91966	AF144378
Brachypodium distachyon	XM_003573939	194033128:54293-55723	133917479
Carica papaya	Phytozome: 162.24_CDS	EU431223:58728-60155	EU431223:2266-3786
Cicer arietinum	XM_004495508	197294093:5003-6430	197294093:2070-3599
Cucumis sativus	XM_004142526	DQ865976:57578-59005	68164782:1838-3376
Fragaria vesca	XM_004304718	325126844:56459-57886	AF288102
Glycine max	XM_003536095; XR137658	91214122:5312-6739	AF142700
Gossypium raimondii	Phytozome:013G120100.1_CDS	372290914:58642-60081	AF403559
Hordeum vulgare	AK353664	AY137453:111-1550	AB078139
Manihot esculenta	Phytozome:03614:2579552..2581338	169794052:58063-59496	EU117376:2063-3583
Medicago truncatula	BT141443	JX512024:117295-118722	AY386945
Nicotiana tobaccum	current study	NC_001879	81238323:2131-3660
Oryza sativa	115482237	AY522330:54082-55536	EU434287
Phaseolus vulgaris	KF033821	EU196765:70304-71734	AY582987
Populus trichocarpa	XM_002319615	134093177:55716-57143	134093177:1981-3513
Ricinus communis	XM_002521916	372450118:58961-60388	372450118:2387-3907
Setaria italica	XM_004982939	558603649:54628-56034	390607728
Solanum lycopersicum	XM004249865	544163592:56683-58116	544163592:2124-3653
Solanum tuberosum	565368659	DQ386163.2|:56531-57964	JF772171:2140-3669
Sorghum bicolor	XM_002448739	118614470:57693-59123	AF164418
Theobroma cacao	Phytozome: EG026242t1_CDS	JQ228389:59398-60852	AY321195
Triticum aestivum	AK334642	AY328025:60-1493	KJ592713:1678-3216
Vitis vinifera	FQ395584; FQ393164	91983971:59436-60863	91983971:2016-3524
Zea mays	226508017	11994090:56874-58304	11994090:1674-3215
Bryophyta			
Pohlia nutans		AY631193	AY522574
Green Algae			
Coccomyxa subellipsoidea	XM_005643171	HQ693844:164006-165433	323149147:70601-72805
Chlorella variabilis	XM_005847023	331268093:47431-48858	331268093:26130-28334
Micromonas pusilla	XM_003063100	FJ858267:20006-21433	FJ858269

Organism	***raf1***	***rbcL***

β-Cyanobacteria

Acaryochloris marina MBIC11017	CP000828 :1771175-1772245	CP000828:1775408-1776838
Anabaena cylindrica PCC 7122	CP003659 :5732014-5733099	CP003659:34579-36009
Anabaena sp 90	CP003284 :2564028-2565113	CP003284:1480330-1481760
Anabaena variabilis ATCC 29413	CP000117 :1756144-1757229	CP000117:4857469-4858899
Calothrix sp PCC 6303	CP003610 :4364743-4365828	CP003610:3605242-3606672
Calothrix sp PCC 7507	CP003943 :5400132-5401217	CP003943:325257-326687
Chamaesiphon minutus PCC 6605	CP003600 :6052812-6053882	CP003600:694685-696115
Chroococcidiopsis thermalis PCC 7203	CP003597 :1959990-1961051	CP003597:5964292-5965722
Crinalium epipsammum PCC 9333	CP003620 :4318634-4319728	CP003620:4709290-4710720
Cyanobacterium aponinum PCC 10605	CP003947 :3620023-3621099	CP003947:800936-802342
Cyanobacterium stanieri PCC 7202	CP003940 :251659-252741	CP003940:126365-127771
Cyanothece sp ATCC 51142	CP000806 :1951795-1952787	CP000806:3281510-3282925
Cyanothece sp PCC 7424	CP001291 :3045110-3046189	CP001291:1503225-1504643
Cyanothece sp PCC 7425	CP001344 :4048780-4049862	CP001344:3372918-3374348
Cyanothece sp PCC 7822	CP002198 :3872031-3873092	CP002198:3223935-3225353
Cyanothece sp PCC 8801	CP001287 :819957-821021	CP001287:1677472-1678890
Cyanothece sp PCC 8802	CP001701 :819755-820819	CP001701:1666285-1667703
Cylindrospermum stagnale PCC 7417	CP003642 :6936516-6937604	CP003642:2391125-2392555
Dactylococcopsis salina PCC 8305	CP003944 :2505154-2506221	CP003944:1798755-1800176
Gloeobacter kilaueensis JS1	CP003587 :711901-712965	CP003587:713821-715245
Gloeobacter violaceus PCC 7421	37508091 :2309302-2310369	37508091:2307046-2308470
Gloeocapsa sp PCC 7428	CP003646 :1785908-1786993	CP003646:1141494-1142924
Halothece sp PCC 7418	CP003945 :2360587-2361660	CP003945:3829408-3830826
Leptolyngbya sp PCC 7376	CP003946 :2022725-2023804	CP003946:204758-206173
Microcoleus sp PCC 7113	CP003630 :771030-772124	CP003630:2675003-2676433
Microcystis aeruginosa PCC 7806	159027328 :13224-14216	166085114:4390428-4391843
Nostoc azollae 708	CP002059 :4390613-4391698	CP002059:2235547-2236977
Nostoc punctiforme PCC 73102	CP001037 :5521656-5522744	CP001037:5263600-5265030
Nostoc sp PCC 7107	CP003548 :2972009-2973094	CP003548:2119530-2120960
Nostoc sp PCC 7120	47118302 :6264560-6265645	47118302:1785970-1787400
Nostoc sp PCC 7524	CP003552 :4087403-4088488	CP003552:1290272-1291702
Oscillatoria acuminata PCC 6304	CP003607 :7273598-7274692	CP003607:1163939-1165369
Oscillatoria nigro-viridis PCC 7112	CP003614 :6651808-6652902	CP003614:6951541-6952971
Pleurocapsa sp PCC 7327	CP003590 :3516618-3517697	CP003590:357448-358863
Pseudanabaena sp PCC 7367	CP003592 :182052-183158	CP003592:1184484-1185896
Rivularia sp PCC 7116	CP003549 :6792297-6793388	CP003549:4304946-4306376
Stanieria cyanosphaera PCC 7437	CP003653 :1606913-1607992	CP003653:369045-370463
Synechococcus elongatus PCC 6301	56684969 :792692-793771	56684969:139920-141338
Synechococcus elongatus PCC 7942	CP000100 :827112-828182	CP000100:1479461-1480879
Synechococcus sp JA-2-3Ba(2-13)	CP000240 :535600-536703	CP000240:2682338-2683762
Synechococcus sp JA-3-3Ab	CP000239 :929252-930337	CP000239:1207204-1208628
Synechococcus sp PCC 6312	CP003558 :1545379-1546446	CP003558:1977136-1978563
Synechococcus sp PCC 7002	CP000951 :2467879-2468958	CP000951:1882749-1884164
Synechococcus sp PCC 7502	CP003594 :3509019-3510092	CP003594:1660201-1661631
Synechocystis sp PCC 6803	359276570 :2974914-2975990	359276570:2476240-2477652
Thermosynechococcus elongatus BP-1	47118315 :1848819-1849889	47118315:1574633-1576060

Figure S1

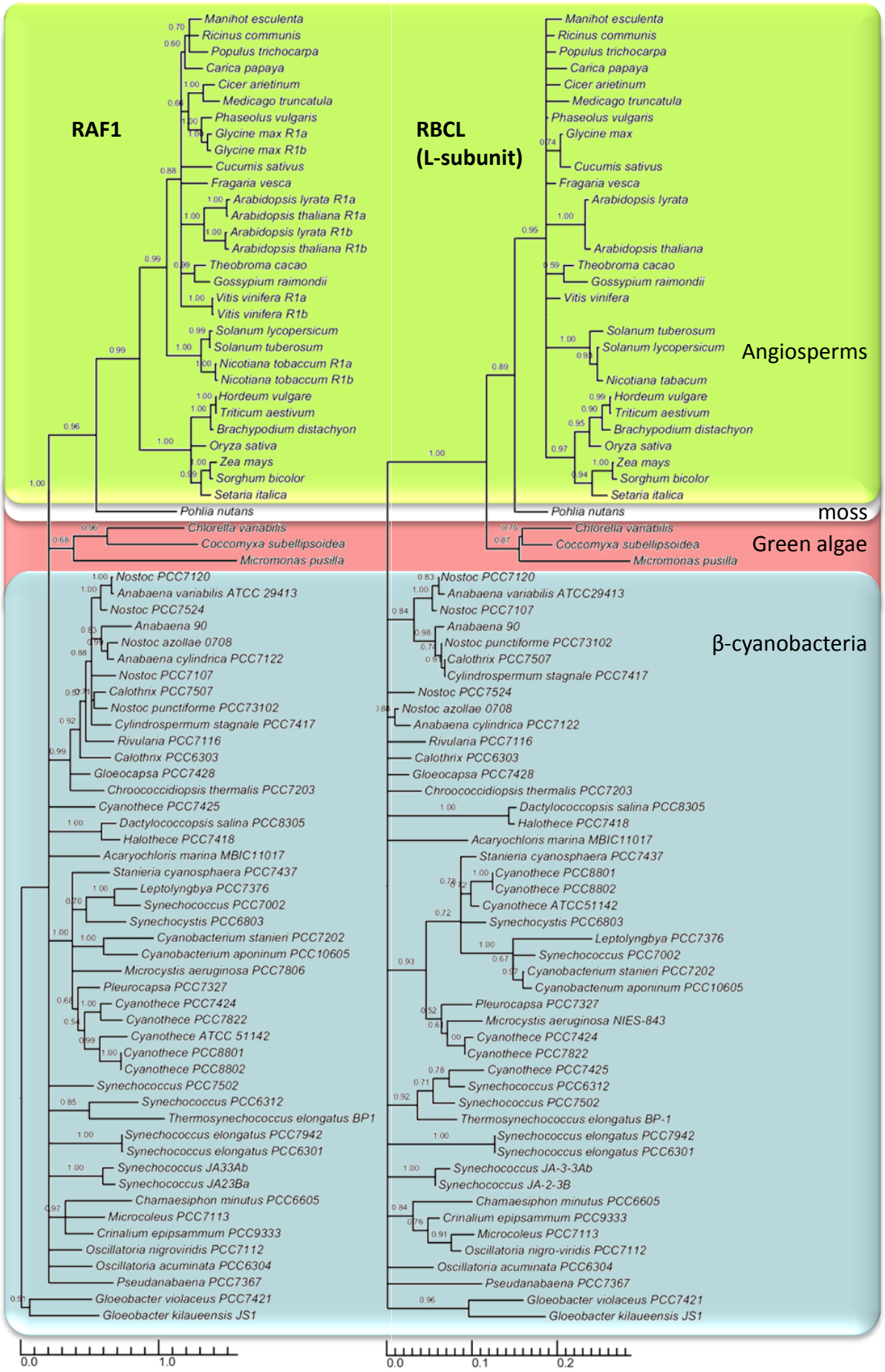

RAF1
RBCL
(L-subunit)
Angiosperms
moss
Green algae
β-cyanobacteria
Manihot esculenta
Ricinus communis
Populus trichocarpa
Carica papaya
Cicer arietinum
Medicago truncatula
Phaseolus vulgaris
Glycine max R1a
Glycine max R1b
Glycine max
Cucumis sativus
Fragaria vesca
Arabidopsis lyrata R1a
Arabidopsis thaliana R1a
Arabidopsis lyrata R1b
Arabidopsis thaliana R1b
Arabidopsis lyrata
Arabidopsis thaliana
Theobroma cacao
Gossypium raimondii
Vitis vinifera R1a
Vitis vinifera R1b
Vitis vinifera
Solanum lycopersicum
Solanum tuberosum
Nicotiana tobaccum R1a
Nicotiana tobaccum R1b
Nicotiana tabacum
Hordeum vulgare
Triticum aestivum
Brachypodium distachyon
Oryza sativa
Zea mays
Sorghum bicolor
Setaria italica
Pohlia nutans
Chlorella variabilis
Coccomyxa subellipsoidea
Micromonas pusilla
Nostoc PCC7120
Anabaena variabilis ATCC 29413
Nostoc PCC7524
Anabaena 90
Nostoc azollae 0708
Anabaena cylindrica PCC7122
Nostoc PCC7107
Calothrix PCC7507
Nostoc punctiforme PCC73102
Cylindrospermum stagnale PCC7417
Rivularia PCC7116
Calothrix PCC6303
Gloeocapsa PCC7428
Chroococcidiopsis thermalis PCC7203
Cyanothece PCC7425
Dactylococcopsis salina PCC8305
Halothece PCC7418
Acaryochloris marina MBIC11017
Stanieria cyanosphaera PCC7437
Leptolyngbya PCC7376
Synechococcus PCC7002
Synechocystis PCC6803
Cyanobacterium stanieri PCC7202
Cyanobacterium aponinum PCC10605
Microcystis aeruginosa PCC7806
Pleurocapsa PCC7327
Cyanothece PCC7424
Cyanothece PCC7822
Cyanothece ATCC 51142
Cyanothece PCC8801
Cyanothece PCC8802
Synechococcus PCC7502
Synechococcus PCC6312
Thermosynechococcus elongatus BP1
Synechococcus elongatus PCC7942
Synechococcus elongatus PCC6301
Synechococcus JA33Ab
Synechococcus JA23Ba
Chamaesiphon minutus PCC6605
Microcoleus PCC7113
Crinalium epipsammum PCC9333
Oscillatoria nigroviridis PCC7112
Oscillatoria acuminata PCC6304
Pseudanabaena PCC7367
Gloeobacter violaceus PCC7421
Gloeobacter kilaueensis JS1
Cyanothece PCC8801
Cyanothece PCC8802
Cyanothece ATCC51142
Synechocystis PCC6803
Leptolyngbya PCC7376
Synechococcus PCC7002
Cyanobacterium stanieri PCC7202
Cyanobactenum aponinum PCC10605
Pleurocapsa PCC7327
Microcystis aeruginosa NIES-843
Cyanothece PCC7424
Cyanothece PCC7822
Cyanothece PCC7425
Synechococcus PCC6312
Synechococcus PCC7502
Thermosynechococcus elongatus BP-1
Synechococcus elongatus PCC7942
Synechococcus elongatus PCC6301
Synechococcus JA-3-3Ab
Synechococcus JA-2-3B
Chamaesiphon minutus PCC6605
Crinalium epipsammum PCC9333
Microcoleus PCC7113
Oscillatoria nigro-viridis PCC7112
Oscillatoria acuminata PCC6304
Pseudanabaena PCC7367
Gloeobacter violaceus PCC7421
Gloeobacter kilaueensis JS1
0.0
1.0
0.0
0.1
0.2

Figure S2

A Amino acid alignment of tobacco and *Arabidopsis* Rubisco L-subunits

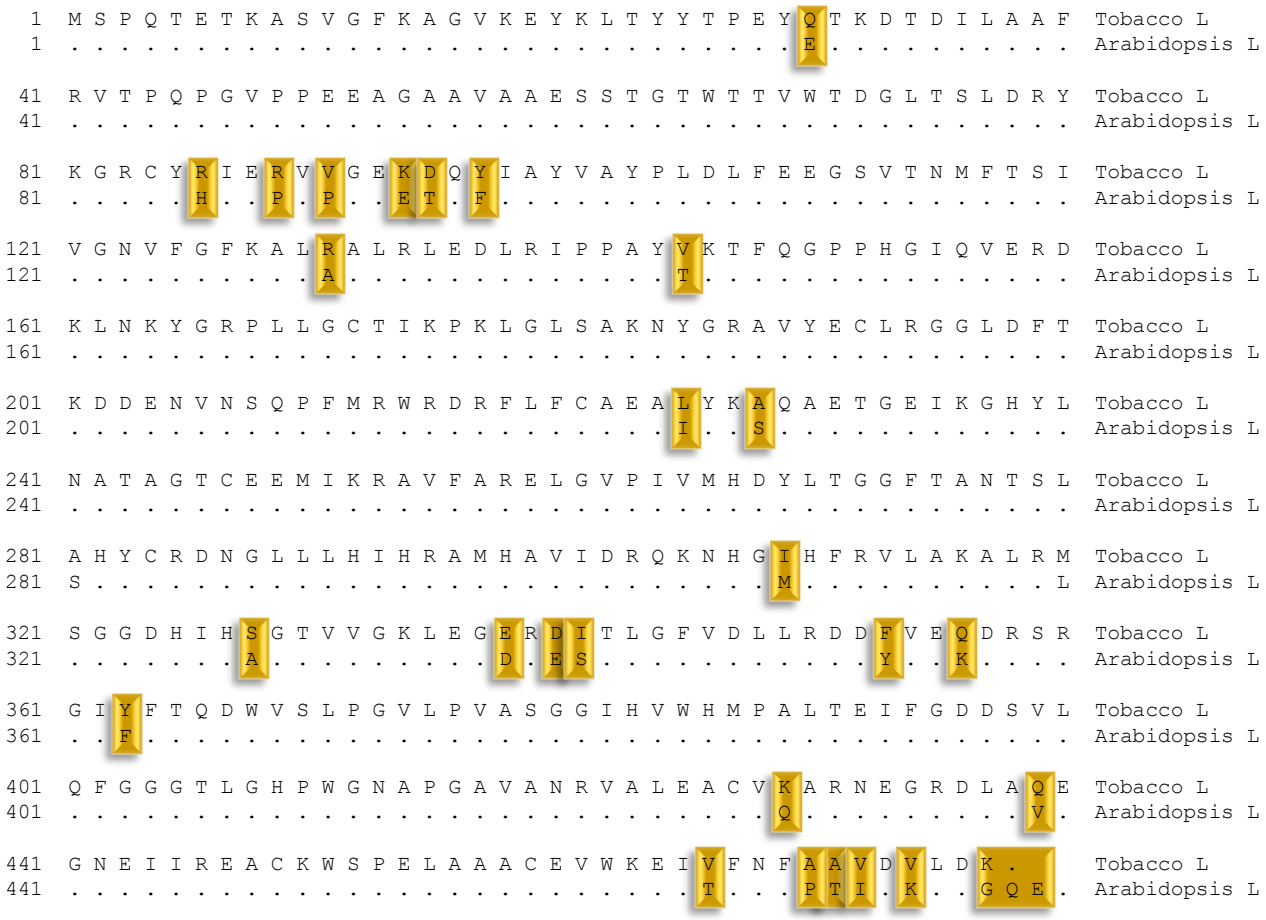
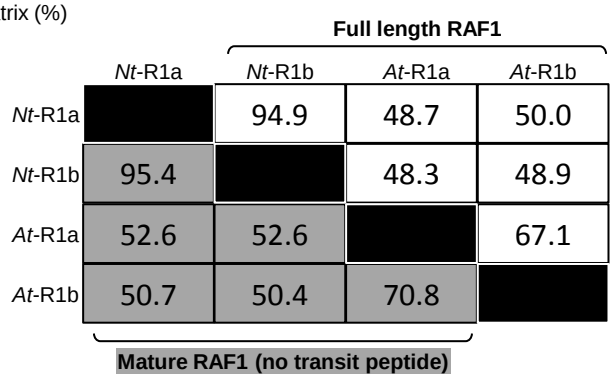


Figure S2

B Amino acid alignment of tobacco (*Nt*) and *Arabidopsis* (*At*) RAF1

C Amino acid sequence identity matrix (%)



Nicotiana tabacum cultivar:K326 Genome sequencing Experiment SRX338110, RNA-Seq transcriptome Illumina analysis (Sierro et al., 2014)

Figure S3

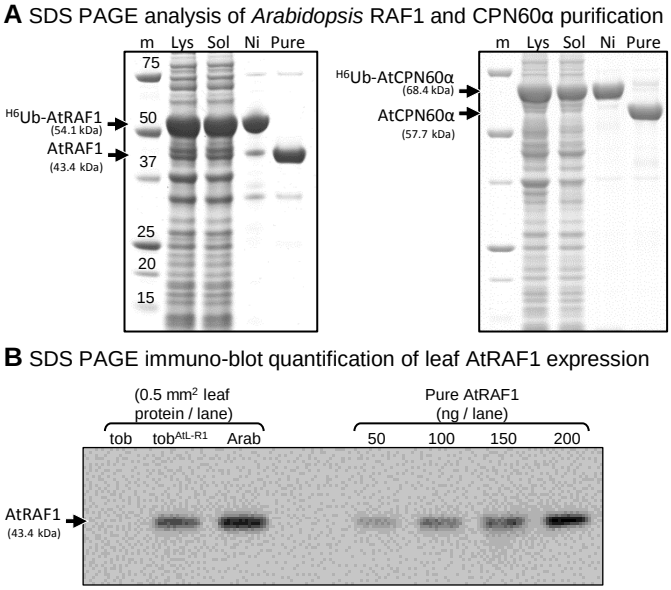

A SDS PAGE analysis of *Arabidopsis* RAF1 and CPN60α purification
m Lys Sol Ni Pure
75
50
37
25
20
15
H6Ub-AtRAF1
(54.1 kDa)
AtRAF1
(43.4 kDa)
m Lys Sol Ni Pure
H6Ub-AtCPN60α
(68.4 kDa)
AtCPN60α
(57.7 kDa)
B SDS PAGE immuno-blot quantification of leaf AtRAF1 expression
(0.5 mm2 leaf
protein / lane)
tob
tobAtL-R1
Arab
Pure AtRAF1
(ng / lane)
50
100
150
200
AtRAF1
(43.4 kDa)

Figure S4

PAGE analysis of leaf soluble and NiNTA purified protein

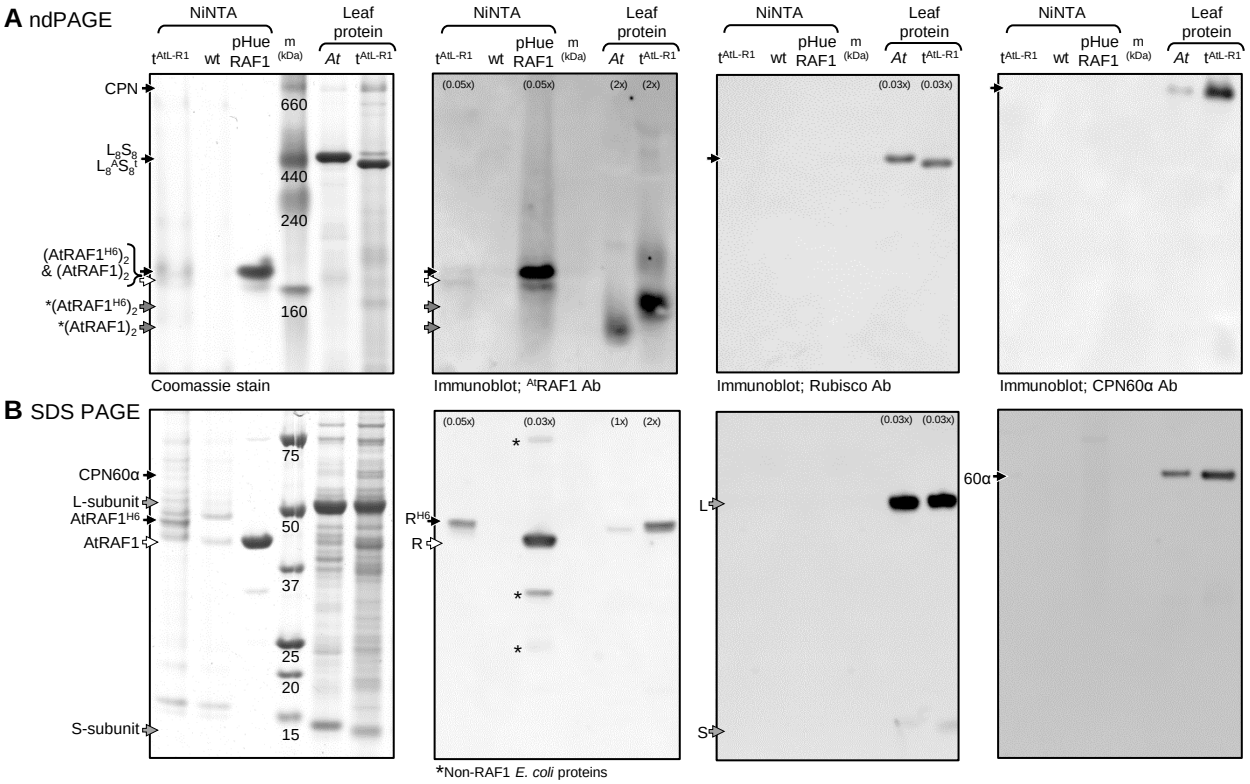
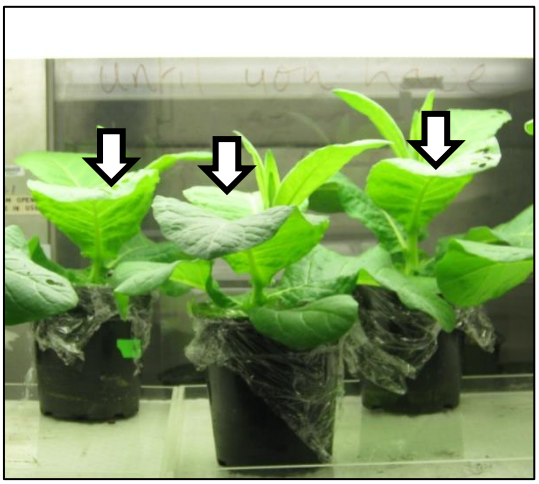


Figure S5

A Plant phenotype and experimental setup for analyzing Rubisco synthesis and turnover in whole leaves by ^{35}S-Met pulse-chase



B Schematic of the leaf pulse-chase analysis abaxial infiltration and sampling régime